

**Amendment and Response**

Applicant: Maria Castellanos et al.

Serial No.: 09/944,919

Filed: August 31, 2001

Docket No.: 10007912-1

Title: METHOD AND SYSTEM FOR MINING A DOCUMENT CONTAINING DIRTY TEXT

---

**REMARKS**

The following remarks are made in response to the Non-Final Office Action mailed September 2, 2003. Claims 1-30 were rejected. With this Response, claims 1, 11, 12, 18 and 21 have been amended and claims 31 and 35 have been added. Claims 13 and 14 have been cancelled. Claims 1-12 and 15-35 remain pending in the application and are presented for consideration and allowance.

**Claim Rejections under 35 U.S.C. § 101**

The Examiner rejected claims 1-10 under 35 U.S.C. § 101 because the claimed invention is directed to non-statutory subject matter. Claim 1 has been amended as suggested by the Examiner to clarify that claims 1-10 are directed to a computer-implemented method. Thus, withdrawal of the rejection under 35 U.S.C. §101 is respectfully requested.

**Claim Rejections under 35 U.S.C. § 102(b)**

The Examiner rejected claims 1-2, 5, 9-12, 15, 19-22, 25, and 29-30 under 35 U.S.C. §102(b) as being anticipated by Domini et al. U.S. Patent No. 6,085,206 (Domini). Applicant submits that the Domini reference fails to disclose the invention of independent claim 1.

Amended independent claim 1 recites a computer-implemented method for mining a document containing dirty text. The method includes removing an instance of dirty text within the document to produce a cleaned document having a content. The method also includes performing a data mining operation on said cleaned document thereby deriving relevant information from said cleaned document and providing a summary of the content of said document.

Domini is directed to removing/correcting dirty text in a document, including correction of both spelling and grammatical construction in a document at the same time. Domini specifically defines dirty text as that text which has not been spell checked and/or that has not been grammar checked (See Domini column 9, lines 43-48). Furthermore, Domini describes that after a sentence has been grammar checked, it is marked as clean text (column 9, lines 49-53).

**Amendment and Response**

Applicant: Maria Castellanos et al.

Serial No.: 09/944,919

Filed: August 31, 2001

Docket No.: 10007912-1

Title: METHOD AND SYSTEM FOR MINING A DOCUMENT CONTAINING DIRTY TEXT

---

Domini fails to describe **performing a data mining operation on a cleaned document**. The Examiner incorrectly concluded that correction of grammar is a form of data mining. Similar to Applicant, Domini defines the extraction of grammatical constructions as removal of dirty text. In further contrast, **Domini fails to disclose performing a data mining operation on the cleaned document thereby deriving relevant information from said cleaned document and providing a summary of the content of said document**. Again, Domini is directed to correcting spelling and grammar in a document, and does not disclose performing a data mining operation that results in providing a summary of the content of the document.

Accordingly, Applicant respectfully requests that the above rejection under 35 U.S.C. § 102(b) should be withdrawn.

Dependant claims 2, 5, 9 and 10 depend directly or indirectly upon independent claim 1. Accordingly, dependant claims 2, 5, 9 and 10 are also allowable over the art of record.

Independent claim 11 is rejected under 35 U.S.C. § 102(b) as being anticipated by Domini. Claims 13 and 14 are rejected under 35 U.S.C. § 103(a) as being unpatentable over Domini in view of U.S. Patent No. 4,965,763 to Zamora (Zamora). Independent claims 11 has been amended to include the limitations of claims 13 and 14, now cancelled.

Accordingly, the rejection of claim 11 is addressed under 35 U.S.C. § 103(a) in light of Domini in view of Zamora.

Claim 11 recites a computer system. The computer system includes a bus, a memory unit coupled to the bus, and a processor coupled to the bus. The processor executes a method for mining a document containing dirty text. The method includes producing a cleaned document having a content including performing a general cleaning of the document by removing an instance of dirty text within the document including instances of misspelling and grammatical errors, performing a domain and task-specific cleaning of the document including removing instances of computer code and tables to produce a cleaned document. A data mining operation is performed on the cleaned document including providing a summary of the content of the document.

For similar reasons as stated above with reference to independent claim 1, Applicant believes independent claim 11 to be allowable over Domini either alone or in view of Zamora.

**Amendment and Response**

Applicant: Maria Castellanos et al.

Serial No.: 09/944,919

Filed: August 31, 2001

Docket No.: 10007912-1

Title: METHOD AND SYSTEM FOR MINING A DOCUMENT CONTAINING DIRTY TEXT

---

Zamora merely recites a computer method for automatic extraction of commonly specified information from business correspondence. The method utilizes a parametric information extraction (PIE) system to identify fields of a business document such the frame slots for a business correspondence or list of business correspondence closing phrases (See Zamora, Fig. 3 and Fig. 5).

Neither Domini nor Zamora teaches or suggests suggests a two-step cleaning process. Specifically, neither Domini nor Zamora teaches or suggests a general cleaning followed by a domain and task-specific cleaning of the document including removing instances of computer code and tables. In view of the above, one could not apply the teachings of Domini, either alone or in combination with Zamora, and arrive at the present invention of independent claim 11.

Dependant claims 12, 15, 19 and 20 depend either directly or indirectly upon independent claim 11. Accordingly, these dependant claims are allowable over the art of record.

Claim 21, recites a computer-useable medium having computer-readable program code embodied therein for causing a computer system to perform removing an instance of dirty text within said document to produce a cleaned document having a content. A data mining operation is performed on said cleaned document to provide a summary of said content. For similar reasons as stated above with reference to independent claim 1, Applicant believes independent claim 21 to be allowable over Domini.

Dependant claims 22, 25, 29 and 30 depend either directly or indirectly upon independent claim 21. Accordingly, these dependant claims are allowable over the art of record.

**Claim Rejections under 35 U.S.C. § 103**

The Examiner rejected claims 3-4, 6-8, 13-14, 16-18, 23-24, and 26-28 under 35 U.S.C. §103(a) as being unpatentable over Domini in view of U.S. Patent No. 4,965,763 to Zamora. Claims 13 and 14 have been cancelled.

All of these claims depend on independent claims 1, 11, or 21, which Applicant believes to be in form for allowance. For the same reasons stated above in reference to

**Amendment and Response**

Applicant: Maria Castellanos et al.

Serial No.: 09/944,919

Filed: August 31, 2001

Docket No.: 10007912-1

Title: METHOD AND SYSTEM FOR MINING A DOCUMENT CONTAINING DIRTY TEXT

---

claims 1, 11, and 21, Applicant believes these corresponding dependent claims to be in allowable form. Withdrawal of the above rejection under 35 U.S.C. § 103 is requested.

**Added Claims**

Applicant has added claims 31-35 directed to a computer-implemented method for mining a document containing dirty text. Applicant believes claims 31-35 to be allowable over the art of record.

**CONCLUSION**

In view of the above, Applicant believes independent claims 1, 11, 21 and 31 and the claims depending therefrom, are in condition for allowance. Allowance of these claims is respectfully requested.

Any inquiry regarding this Amendment and Response should be directed to either Steven E. Dicke at Telephone No. (612) 573-2002, Facsimile No. (612) 573-2005 or Howard Boyle at Telephone No. (281) 518-9645, Facsimile No. (218) 514-8332. In addition, all correspondence should continue to be directed to the following address:

**Hewlett-Packard Company**  
Intellectual Property Administration  
P.O. Box 272400  
Fort Collins, Colorado 80527-2400

**Amendment and Response**

Applicant: Maria Castellanos et al.

Serial No.: 09/944,919

Filed: August 31, 2001

Docket No.: 10007912-1

Title: METHOD AND SYSTEM FOR MINING A DOCUMENT CONTAINING DIRTY TEXT

---

Respectfully submitted,

Maria Castellanos et al.,

By their attorneys,

DICKE, BILLIG & CZAJA, PLLC

Fifth Street Towers, Suite 2250

100 South Fifth Street

Minneapolis, MN 55402

Telephone: (612) 573-2000

Facsimile: (612) 573-2005

Date: December 2, 2003  
SED:jan

Steven E. Dicke  
Steven E. Dicke  
Reg. No. 38,431

**CERTIFICATE UNDER 37 C.F.R. 1.8:** The undersigned hereby certifies that this paper or papers, as described herein, are being deposited in the United States Postal Service, as first class mail, in an envelope address to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on this 2 day of December, 2003.

By Steven E. Dicke  
Name: